Urine Steroid Metabolomics as a Biomarker Tool for Detecting Malignancy in Adrenal Tumors

Arlt, Wiebke; Biehl, Michael; Taylor, Angela E.; Hahner, Stefanie; Libe, Rossella; Hughes, Beverly A.; Schneider, Petra; Smith, David J.; Stiekema, Han; Krone, Nils

Published in:
Journal of Clinical Endocrinology and Metabolism

DOI:
10.1210/jc.2011-1565

Publication date:
2011

Supplemental methods section

Computational analysis of steroid metabolite excretion (supplementary material)

For computational steroid excretion data analysis, we employed a recently developed variant of Learning Vector Quantization (LVQ) [1], Generalized Matrix Relevance LVQ (GMLVQ) [2, 3], For GMLVQ experiments, all numerical steroid excretion values were log-transformed and subsequently normalized by subtracting the respective mean values obtained in healthy controls with a similar age and sex distribution (n=88) and dividing by the corresponding standard deviations. This yielded 32 log-transformed steroid excretion values for each patient, expressed on a scale set by the control group. The vectors of log-transformed excretion values (rescaled with respect to the controls) labeled with the corresponding class membership, i.e. either ACC or ACA, served as input for the machine learning system.

For evaluation of the performance of the trained system, we split the ACC and ACA data into a training and test set, respectively, with 90% of the data from each class used in the training process to determine prototypes and relevance matrix. The remaining 10% served as a test set, which was used to quantify the true positive rate, i.e. the fraction of ACC test data classified correctly as ACC, and the false positive rate, i.e. the fraction of ACA test data misclassified by the system as ACC. In order to suppress the influence of possible *lucky* or *unlucky draws* of the test set on the performance of the system, we repeated the procedure for 1000 randomized training/test data set compositions. As an additional preprocessing step, we applied a z-score transformation in each training process, such that for all markers the mean of the transformed excretion over the training set was *zero* and the corresponding variance was *one*. We employed a single prototype per class. Initial values for the prototypes were determined from the class conditional means of a random selection of training data. The relevance matrix was initially set as the identity, i.e. all steroids equally important and pair-wise relevance being zero.

Subsequently, we obtained the full Receiver Operating Characteristics (ROC) of the classifier, generated by varying the threshold generalized distance margin separating the two classes, and computed the area under curve (AUC). The mean ROC and standard deviations were calculated using threshold averaging, as described in [4].

We have applied the GMLVQ machine learning analysis to the complete panel of 32 steroid markers. In addition, we studied the selection of discriminant features in GMLVQ: In each training run we determined the subsets of the most relevant three or nine steroids from the obtained relevance matrix. The GMLVQ training process was then repeated, restricting the system to the use of the individually selected subsets of steroids. The resulting mean ROC curves quantify the achievable classification performance when using reduced panels of three or nine steroids selected by GMLVQ.

For comparison, we also employed statistical modeling techniques, Fisher Linear Discriminant Analysis (LDA) and a standard implementation of logistic regression (27). Logistic

Regression was implemented using the NAG® library routine G02GB (Numerical Algorithms Group, Oxford, UK), for LDA we employed the implementation in van der Maaten's MATLAB® Toolbox for dimensionality reduction.

For a small number of data points, the excretion values were found to be zero within the sensitivity of the GC/MS analysis; for log-transformation and machine learning analysis, these values were set to 10^{-10}. It was verified that the choice of this correction parameter did not affect the classification. While steroid data in ACA patients (n=102) displayed no missing values, the available steroid excretion data for ACC patients (n=45) contained a total of 3.9% (56 out of 1440) missing steroid excretion values. For the LVQ analysis, missing values were ignored when comparing distances of a particular feature vector from different prototypes. For LDA and logistic regression analyses we replaced missing values by class conditional means, which should theoretically give a performance advantage to these two methods in comparison to GMLVQ.

References

1. Kohonen, T., Self-organizing maps. 2nd ed. Springer series in information sciences, 300720-678X. 1997, Berlin ; New York: Springer. xvii, 426 p.
2. Schneider, P., M. Biehl, and B. Hammer, Adaptive Relevance Matrices in Learning Vector Quantization. Neural Computation, 2009. **21**(12): p. 3532-3561.
3. Schneider, P., M. Biehl, and B. Hammer, Distance Learning in Discriminative Vector Quantization. Neural Computation, 2009. **21**(10): p. 2942-2969.
4. Fawcett, T., An introduction to ROC analysis. Pattern Recognition Letters, 2006. **27**(8): p. 861-874.
5. Duda, R.O., P.E. Hart, and D.G. Stork, Pattern classification. 2nd ed. 2000, New York: Wiley. xx, 654p.
6. Cox, D.R. and E.J. Snell, Analysis of binary data. 2nd ed. Monographs on statistics and applied probability. 1989, London: Chapman and Hall. xi, 236 p.
7. P. McCullagh and J.A. Nelder. Generalized linear models. Chapman & Hall/CRC, 1989.
8. van der Maaten, L.J.P., Matlab Toolbox for Dimensionality Reduction (v0.7b). URL: http://homepage.tudelft.nl/19j49/Matlab_Toolbox_for_Dimensionality_Reduction.html, published online 2007.

Suppl. Table 1. Steroids and steroid metabolites analyzed in a single diagnostic run by gas chromatography/mass spectrometry (GC/MS).

No.	Abbreviation	Common name	Chemical name	Metabolite of
Androgen metabolites				
1	An	Androsterone	5α-androstan-3α-ol-17-one	Androstenedione, testosterone, 5α-dihydrotestosterone
2	Etio	Etiocholanolone	5β-androstan-3α-ol-17-one	Androstenedione, testosterone
Androgen precursor metabolites				
3	DHEA	Dehydroepi-androsterone	5-androsten-3β-ol-17-one	DHEA + DHEA sulfate (DHEAS)
4	16α-OH-DHEA	16α-hydroxy-DHEA	5-androstene-3β,16α-diol-17-one	DHEA + DHEAS
5	5-PT	Pregnenetriol	5-pregnene-3β,17, 20α-triol	17-hydroxypregnenolone
6	5-PD	Pregnenediol	5-pregnene-3β, 20α-diol and 5, 17, (20)-pregnadien-3β-ol	pregnenolone
Mineralocorticoid metabolites				
7	THA	Tetrahydro-11-dehydro-corticosterone	5β-pregnane-3α, 21-diol, 11, 20-dione	corticosterone, 11-dehydrocorticosterone
8	5α-THA	5α-tetrahydro-11-dehydro-corticosterone	5α-pregnane-3α, 21-diol-11, 20-dione	corticosterone, 11-dehydrocorticosterone
9	THB	Tetrahydro-corticosterone	5β-pregnane-3α, 11β, 21-triol-20-one	corticosterone
10	5α-THB	5α-tetrahydro-corticosterone	5α-pregnane-3α, 11β, 21-triol-20-one	corticosterone
11	3α5β-THALDO	Tetrahydro-aldosterone	5β-pregnane-3α, 11β, 21-triol-20-one-18-al	aldosterone
Mineralocorticoid precursor metabolites				
12	THDOC	Tetrahydro-11-deoxycorticosterone	5β-pregnane-3α, 21-diol-20-one	11-deoxycorticosterone
13	5α-THDOC	5α-tetrahydro-11-deoxycorticosterone	5α-pregnane-3α, 21-diol-20-one	11-deoxycorticosterone

Glucocorticoid precursor metabolites				
14	PD	Pregnanediol	5β-pregnane-3α, 20a-diol	progesterone
15	3α5α-17HP	3α, 5α-17-hydroxy-pregnanolone	5α-pregnane-3α, 17α-diol-20-one	17-hydroxyprogesterone
16	17HP	17-hydroxy-pregnanolone	5β-pregnane-3α, 17α,-diol-20-one	17-hydroxyprogesterone
17	PT	Pregnanetriol	5β-pregnane-3α, 17α, 20α-triol	17-hydroxyprogesterone
18	PTONE	Pregnanetriolone	5β-pregnane-3α, 17, 20α-triol-11-one	21-deoxycortisol
19	THS	Tetrahydro-11-deoxycortisol	5β-pregnane-3α, 17, 21-triol-20-one	11-deoxycortisol
Glucocorticoid metabolites				
20	F	Cortisol	4-pregnene-11β, 17, 21-triol-3, 20-dione	cortisol
21	6β-OH-F	6β-hydroxy-cortisol	4-pregnene-6β, 11β, 17, 21-tetrol-3, 20-dione	cortisol
22	THF	Tetrahydrocortisol	5β-pregnane-3α, 11β, 17, 21-tetrol-20-one	cortisol
23	5α-THF	5α-tetrahydrocortisol	5α-pregnane-3α, 11β, 17, 21-tetrol-20-one	cortisol
24	α-cortol	α-cortol	5β-pregnan-3α, 11β, 17, 20α, 21-pentol	cortisol
25	β-cortol	β-cortol	5β-pregnan-3α, 11β, 17, 20β, 21-pentol	cortisol
26	11β-OH-An	11β-hydroxy-androsterone	5α-androstane-3α, 11β-diol-17-one	cortisol (+androgens)
27	11β-OH-Et	11β-hydroxy-etiocholanolone	5β-androstane-3α, 11β-diol-17-one	cortisol (+androgens)
28	E	Cortisone	4-pregnene-17α, 21-diol-3, 11, 20-trione	cortisone
29	THE	Tetrahydrocortisone	5β-pregnene-3α, 17, 21-triol-11, 20-dione	cortisone
30	α-cortolone	α-cortolone	5β-pregnane-3α, 17, 20α, 21-tetrol-11-one	cortisone
31	β-cortolone	β-cortolone	5β-pregnane-3α, 17, 20β, 21-tetrol-11-one	cortisone
32	11-oxo-Et	11-oxo-etiocholanolone	5β-androstan-3α-ol-11, 17-dione	cortisone (+androgens)

Suppl. Table 2. Urinary excretion of steroid metabolites (median (interquartile range, IQR)) in healthy controls, patients with adrenocortical adenomas (ACA), and adrenocortical carcinoma patients (ACC). Steroids are numbered as in Fig. 1. Statistical analysis was performed employing Kruskal-Wallis nonparametric testing and Dunn's post hoc test.

Steroid number	Urinary steroid excretion (µg/24h)	Controls (n=88)	ACA (n=102)	ACC (n=45)
1	An Median (IQR)	1258 (814-2032)	632 (302-1117) $p=8x10^{-9}$*	1130 (515-2445) p=0.73* $p=8x10^{-4}$†
2	Etio	1547 (915-2231)	803 (402-1197) $p=1x10^{-8}$	3671 (1171-7372) p=0.031 $p=3x10^{-13}$
3	DHEA	165 (80-460)	58 (21-135) $p=5x10^{-6}$	612 (98-18273) p=0.058 $p=9x10^{-10}$
4	16α-DHEA	278 (186-728)	201 (87-366) p=0.036	653 (169-3168) p=0.30 $p=6x10^{-7}$
5	5-PT	195 (127-307)	121 (78-170) $p=3x10^{-4}$	1901 (554-7865) $p=9x10^{-8}$ $p=4x10^{-18}$
6	5-PD	168 (104-247)	257 (140-374) p=0.15	3128 (848-14308) $p=7x10^{-12}$ $p=3x10^{-12}$
7	THA	123 (73-190)	94 (62-173) p=0.26	112 (56-195) p=0.26 p=0.26
8	5α-THA	113 (63-177)	88 (64-132) p=0.076	76 (42-160) p=0.076 p=0.076
9	THB	124 (79-181)	105 (68-171) p=0.19	147 (61-380) p=0.19 p=0.19

10	5α-THB	225 (143-387)	221 (145-330) p=0.19	155 (77-356) p=0.19 p=0.19
11	3α5β- THALDO	41 (29-57)	22 (14-34) p=0.034	24 (11-48) p=0.18 p=1.0
12	THDOC	23 (10-40)	16 (10-27) p=1.0	103 (36-222) $p=7x10^{-4}$ $p=4x10^{-11}$
13	5α-THDOC	3 (2-7)	4 (2-7) p=1.0	22 (7-66) $p=4x10^{-5}$ $p=2x10^{-9}$
14	PD	197 (122-353)	138 (85-206) p=0.001	839 (364-2657) $p=3x10^{-6}$ $p=7x10^{-15}$
15	3α5α-17HP	9 (5-19)	9 (4-21) p=1.0	19 (11-44) p=0.085 p=0.002
16	17HP	139 (68-288)	120 (56-229) p=0.47	511 (299-1021) $p=2x10^{-9}$ $p=2x10^{-13}$
17	PT	473 (269-728)	373 (209-724) p=0.74	1484 (769-4199) $p=2x10^{-9}$ $p=1x10^{-12}$
18	PTONE	13 (7-22)	18 (10-35) p=0.012	32 (10-99) p=0.001 p=0.50
19	THS	47 (33-63)	122 (84-189) $p=1x10^{-14}$	2151 (334-4703) $p=2x10^{-30}$ $p=1x10^{-7}$
20	F	66 (44-105)	85 (52-127) p=0.048	245 (95-701) $p=8x10^{-10}$ $p=2x10^{-5}$
21	6β-OH cortisol	100 (53-129)	133 (76-205) p=0.18	356 (187-1530) $p=2x10^{-6}$ $p=3x10^{-6}$

22	THF	1173 (840-1820)	1811 (1261-2433) $p=2x10^{-5}$	2836 (1373-5234) $p=7x10^{-9}$ p=0.04
23	5α-THF	817 (502-1221)	1264 (692-2115) $p=7x10^{-4}$	852 (450-1711) p=1.0 p=0.11
24	α-cortol	227 (159-323)	355 (243-517) $p=1x10^{-5}$	557 (279-1589) $p=1x10^{-10}$ p=0.008
25	β-cortol	246 (182-384)	536 (345-764) $p=5x10^{-9}$	740 (316-1341) $p=3x10^{-9}$ p=0.50
26	11β-OH-An	473 (336-843)	552 (353-872) p=0.24	623 (346-1921) p=0.24 p=0.24
27	11β-OH-Et	257 (136-401)	265 (135-453) p=0.18	366 (149-1668) p=0.18 p=0.18
28	E	120 (78-193)	126 (74-175) p=1.0	164 (87-364) p=0.088 p=0.061
29	THE	2347 (1535-3285)	3478 (2185-4779) $p=4x10^{-5}$	3701 (1909-6852) $p=2x10^{-4}$ p=1.0
30	α-cortolone	883 (690-1266)	1340 (1013-1802) $p=3x10^{-5}$	1840 (868-2853) $p=5x10^{-5}$ p=1.0
31	β-cortolone	424 (306-642)	666 (436-957) $p=3x10^{-5}$	677 (458-1312) $p=4x10^{-4}$ p=1.0
32	11-oxo-Et	363 (207-539)	401 (204-667) p=0.11	484 (232-1903) p=0.11 p=0.11

* Comparison of Controls vs. ACA and ACC, respectively

† Comparison of ACA vs. ACC

Suupl. Table 3. Urinary excretion of steroid metabolites by steroid subclass (median (interquartile range, IQR)) in ACA and ACC patients with or without evidence of hormone excess (according to the results of the routine biochemistry, see methods section for a detailed description) and in healthy controls.

Urinary excretion of steroid metabolites subclasses (µg/24h)	ACA without hormone excess (n=69)	ACA with hormone excess (n=33)	Healthy controls (n=88)	ACC without hormone excess (n=12)	ACC with hormone excess (n=33)
Androgen metabolites	1254 (698-1849)	2292 (914-3128)	2787 (1808-4305)	1866 (888-3801)	6953 (2290-10583)
Median (IQR)	$p=2x10^{-11}$*	p=0.13*		p=0.87*	p=0.023 *
	p=0.032†			p=0.007†	
Androgen precursor	592 (352-1051)	806 (634-1587)	493 (320-944)	1872 (421-8809)	10287 (4684-102057)
metabolites	p=1.0	p=0.052		p=0.015	$p=8x10^{-18}$
Median (IQR)	p=0.34			p=0.14	
Mineralocorticoid	535 (378-811)	610 (400-1256)	598 (355-941)	596 (433-836)	692 (281-1196)
metabolites	p=0.83	p=0.83		p=0.83	p=0.83
Median (IQR)	p=0.83			p=0.83	
Mineralocorticoid	18 (12-28)	25 (16-47)	25 (13-44)	39 (31-99)	219 (72-408)
precursor metabolites	p=1.0	p=1.0		p=0.66	$p=3x10^{-5}$
Median (IQR)	p=0.35			p=0.47	
Glucocorticoid precursor	750 (468-1289)	1043 (754-1797)	973 (570-1317)	2147 (1510-5681)	10599 (4823-30083)
metabolites	p=1.0	p=1.0		p=0.002	$p=3x10^{-15}$
Median (IQR)	p=0.11			p=1.0	
Glucocorticoid	11642 (8064-14561)	11716 (9480-20065)	7763 (5639-11382)	9208 (5581-11682)	20875 (11898-46110)
metabolites	$p=6x10^{-4}$	$p=8x10^{-4}$		p=1.0	$p=3x10^{-10}$
Median (IQR)	p=1.0			p=0.008	

* comparison to healthy controls

†comparison between patient subgroups with and without evidence of hormone excess

Suppl. Fig. 1

Heat map showing individual excretion value for the 32 steroid metabolites in the 102 ACA and 45 ACC patients. The color code represents the log-transformed steroid excretion values that were normalized to the healthy controls (n=88) by subtracting the respective mean values obtained in the controls and dividing by the corresponding standard deviation.